



DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIÈRE DE BREVETS (PCT)

(51) Classification internationale des brevets ⁶ : G06F 15/80	A1	(11) Numéro de publication internationale: WO 99/01825 (43) Date de publication internationale: 14 janvier 1999 (14.01.99)
--	-----------	--

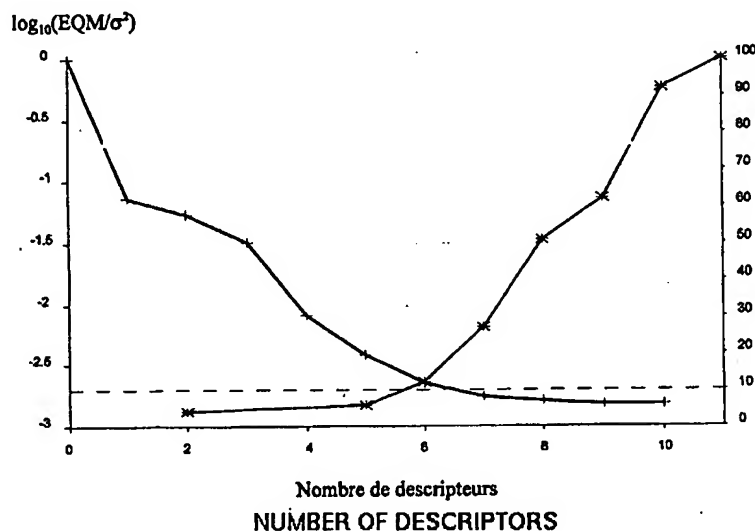
<p>(21) Numéro de la demande internationale: PCT/FR98/01381</p> <p>(22) Date de dépôt international: 29 juin 1998 (29.06.98)</p> <p>(30) Données relatives à la priorité: 97/08342 2 juillet 1997 (02.07.97) FR</p> <p>(71) Déposant (pour tous les Etats désignés sauf US): INFORMATIQUE CDC [FR/FR]; 56, rue de Lille, F-75007 Paris (FR).</p> <p>(72) Inventeurs; et (75) Inventeurs/Déposants (US seulement): STOPPIGLIA, Hervé [FR/FR]; 44, chemin des Prés, F-73000 Barberaz (FR). DREYFUS, Gérard [FR/FR]; 16, rue de la Source Perdue, F-91190 Gif-sur-Yvette (FR).</p> <p>(74) Mandataire: ARNAUD, Jean; Cabinet Amaud, 94, rue Saint-Lazare, F-75009 Paris (FR).</p>	<p>(81) Etats désignés: CA, CN, JP, US, brevet européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</p> <p>Publiée Avec rapport de recherche internationale.</p>
--	--

(54) Title: METHOD FOR CONSTRUCTING A NEURAL NETWORK FOR MODELLING A PHENOMENON

(54) Titre: PROCEDE DE CONSTRUCTION D'UN RESEAU DE NEURONES POUR LA MODELISATION D'UN PHENOMENE

(57) Abstract

The invention concerns the construction of a neural network, more particularly it concerns a method consisting in determining variables to be incorporated in an optimal model by evaluating results, and the construction of a neural network by determining the neural links based on a resulting model. It consists in inputting an additional variable which has random values, determining and classifying descriptors of the variables, by applying a criterion comparing the results, for determining an order of decreasing significance of the descriptors, then eliminating a variable whereof the descriptor is ranked after that of the additional variable. The invention is useful for modelling phenomena.



(57) Abrégé

L'invention concerne la construction d'un réseau de neurones. Elle se rapporte à un procédé qui comprend la détermination des variables à incorporer au modèle optimal par évaluation des résultats, et la construction d'un réseau de neurones par détermination des liaisons des neurones en fonction d'un modèle obtenu. Il comprend l'introduction d'une variable supplémentaire qui possède des valeurs aléatoires, la détermination et le classement des descripteurs des variables, par application d'un critère de comparaison des résultats, pour la détermination d'un ordre de signification décroissante des descripteurs, puis l'élimination d'une variable dont le descripteur est classé après celui de la variable supplémentaire. Application à la modélisation des phénomènes.

UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	ES	Espagne	LS	Lesotho	SI	Slovénie
AM	Arménie	FI	Finlande	LT	Lituanie	SK	Slovaquie
AT	Autriche	FR	France	LU	Luxembourg	SN	Sénégal
AU	Australie	GA	Gabon	LV	Lettonie	SZ	Swaziland
AZ	Azerbaïdjan	GB	Royaume-Uni	MC	Monaco	TD	Tchad
BA	Bosnie-Herzégovine	GE	Géorgie	MD	République de Moldova	TG	Togo
BB	Barbade	GH	Ghana	MG	Madagascar	TJ	Tadjikistan
BE	Belgique	GN	Guinée	MK	Ex-République yougoslave de Macédoine	TM	Turkménistan
BF	Burkina Faso	GR	Grèce	ML	Mali	TR	Turquie
BG	Bulgarie	HU	Hongrie	MN	Mongolie	TT	Trinité-et-Tobago
BJ	Bénin	IE	Irlande	MR	Mauritanie	UA	Ukraine
BR	Brésil	IL	Israël	MW	Malawi	UG	Ouganda
BY	Bélarus	IS	Islande	MX	Mexique	US	Etats-Unis d'Amérique
CA	Canada	IT	Italie	NE	Niger	UZ	Ouzbékistan
CF	République centrafricaine	JP	Japon	NL	Pays-Bas	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norvège	YU	Yougoslavie
CH	Suisse	KG	Kirghizistan	NZ	Nouvelle-Zélande	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	République populaire démocratique de Corée	PL	Pologne		
CM	Cameroun	KR	République de Corée	PT	Portugal		
CN	Chine	KZ	Kazakstan	RO	Roumanie		
CU	Cuba	LC	Sainte-Lucie	RU	Fédération de Russie		
CZ	République tchèque	LI	Liechtenstein	SD	Soudan		
DE	Allemagne	LK	Sri Lanka	SE	Suède		
DK	Danemark	LR	Libéria	SG	Singapour		
EE	Estonie						

Procédé de construction d'un réseau de neurones pour la modélisation d'un phénomène

La présente invention concerne un procédé de construction d'un réseau de neurones destiné à la modélisation d'un phénomène, ainsi qu'un réseau de neurones réalisé par mise en oeuvre du procédé selon l'invention.

Bien que l'invention concerne un perfectionnement apporté à la construction des réseaux de neurones, on ne décrit pas l'ensemble des principes et procédés utilisés de façon classique pour la construction des réseaux de neurones car ils sont connus depuis longtemps, et il existe une littérature considérable à leur sujet. On ne décrit donc la technique antérieure que dans la mesure où l'invention s'y rapporte, et on décrit l'invention en référence à ces aspects de la technique antérieure.

On considère d'abord un certain nombre de définitions utilisées dans le présent mémoire et qui sont en général celles de la technique antérieure.

Les réseaux de neurones sont des circuits matériels, réalisés par exemple sous forme de circuits intégrés, mais qui peuvent aussi être réalisés uniquement sous forme d'un logiciel. Un neurone est un élément qui possède des entrées destinées à recevoir des signaux représentatifs de variables et une ou plusieurs sorties, et il transmet des données de sortie ou de résultat par application d'une fonction d'activation.

Dans un réseau de neurones, il existe, en plus des neurones, des entrées, au moins une sortie et des liaisons formées entre les entrées et les neurones et entre les neurones et les sorties, et éventuellement entre les neurones. On démontre qu'un réseau de neurones du type précité et qui comporte plusieurs couches, c'est-à-dire ayant des liaisons en cascade entre des sorties de neurones d'une couche et des entrées de neurones d'une autre couche, est équivalent à un réseau de neurones à une seule couche dite "cachée", c'est-à-dire que tous les neurones du réseau n'ont des liaisons qu'avec les entrées et les sorties.

Les variables sont des grandeurs qui peuvent prendre plusieurs valeurs et participent au phénomène qu'on veut modéliser. Le phénomène qu'on veut modéliser peut être quelconque, mais les réseaux de neurones sont évidemment appliqués à des phénomènes dont on ne connaît pas la fonction qui lie les variables au résultat. Si l'on connaît cette fonction, il est plus simple et plus précis de réaliser directement un circuit mettant en oeuvre la fonction.

Un modèle d'un phénomène est représenté à la fois par l'ensemble des variables et par le traitement qu'elles subissent pour donner le résultat, notamment par les fonctions d'activation des neurones. Un sous-modèle est un modèle dont une variable au moins a été éliminé.

La validité d'un modèle est déterminée par apprentissage, c'est-à-dire par utilisation, comme signaux d'entrée, des valeurs de variables qui ont été déterminées et dont on connaît le résultat. L'apprentissage comprend l'application de plusieurs exemples, c'est-à-dire de plusieurs groupes de valeurs de variables, avec obtention de résultats qui peuvent être comparés aux résultats des exemples.

L'appréciation de la validité d'un modèle s'effectue par comparaison du résultat obtenu dans l'apprentissage avec le résultat de l'exemple considéré.

On utilise aussi des "descripteurs" qui sont des ensembles des valeurs d'une même variable dans un ensemble d'exemples utilisés pour l'apprentissage. Cet ensemble des valeurs de variables peut avoir diverses formes. Dans un exemple particulièrement intéressant, constituant un mode de réalisation préféré de l'invention, les descripteurs sont des vecteurs à N dimensions, N étant le nombre d'exemples utilisés pour l'apprentissage. Ces vecteurs agissent donc dans un espace à N dimensions. Chacun de ces vecteurs est orthogonal à un espace à $N-1$ dimensions qui est défini comme étant l'espace à $N-1$ dimensions dans lequel la projection du vecteur du descripteur, supposé non nul, est un point.

Les définitions qui précèdent des termes utilisés dans le présent mémoire suggèrent déjà le problème auquel

s'applique l'invention et qui est la modélisation d'un phénomène, permettant la réalisation optimale d'un réseau de neurones dont les entrées reçoivent les valeurs des variables et dont la sortie ou les sorties représentent des données de résultat. Le procédé de construction d'un tel réseau de neurones comprend en général, de manière connue, une première phase qui, à partir d'un groupe de variables trop important, détermine les seules variables qui doivent être utilisées parce qu'elles ont une signification dans le phénomène, et une seconde phase de construction d'un réseau de neurones optimal qui, à partir des signaux représentatifs des valeurs des variables, transmet des données de résultat représentant le phénomène.

De manière connue, la première phase comprend la détermination de descripteurs, en nombre excessif, et la sélection, parmi l'ensemble des modèles possibles, de celui qui explique de la meilleure manière possible le phénomène observé. Il faut noter que cette explication doit tenir compte de la performance du modèle (faible écart entre le résultat donné par le modèle et les observations), mais aussi de sa complexité (notamment parce que le traitement doit être aussi rapide que possible).

On pourrait évaluer tous les modèles possibles. Il faut noter qu'un modèle a un type (par exemple linéaire ou non, statique ou dynamique, ...), une structure (définie par la famille de fonctions envisagées et l'ensemble des variables descriptives nécessaires), et des paramètres (qui définissent la fonction choisie parmi la famille F de fonctions). Une première possibilité de sélection d'un modèle comprend la prise en considération d'un modèle complet utilisant tous les descripteurs, puis la réalisation de tous les sous-modèles possibles et, parmi ces sous-modèles possibles, la sélection du meilleur. Il faut alors estimer un nombre de modèles extrêmement important. En effet, lorsque le nombre de variables, et donc de descripteurs, est égal à P , il faut estimer 2^P modèles séparément. Par exemple, lorsque l'ensemble comporte quinze

variables, le nombre de modèles possibles à comparer est de 32 768. Ce nombre devient vite extrêmement grand si bien que ce procédé devient rapidement inutilisable.

On a donc mis au point d'autres procédés permettant de
5 réduire le nombre de modèles à évaluer. On connaît ainsi des
procédés destructif et constructif. Dans le premier procédé,
on utilise, à partir du modèle complet à P descripteurs,
tous les sous-modèles possibles à P-1 descripteurs, on
sélectionne celui qui donne la meilleure performance, et, si
10 le sous-modèle est meilleur que le modèle complet, on
reprend la procédure à partir de celui-ci alors que, s'il
n'est pas meilleur que le modèle complet, on repart du
modèle complet. Dans le procédé "constructif", on part d'un
modèle à 0 descripteur et on construit les P modèles à 1
15 descripteur, on choisit le meilleur de ces modèles et on
poursuit la procédure par addition d'un descripteur, jusqu'à
ce que le modèle obtenu soit meilleur que tous les modèles
obtenus par augmentation du nombre de descripteurs d'une
unité. Ces deux procédés permettent une réduction très
20 importante du nombre de modèles à évaluer.

Par rapport au procédé de sélection de modèle par
évaluation de tous les modèles, les deux procédés précités
peuvent ne pas donner un modèle optimal. Cependant, ils
doivent être souvent utilisés dans la mesure où l'évaluation
25 de la totalité des modèles possibles est en dehors des
possibilités des machines disponibles de calcul. Lorsque les
deux procédés (destructif et constructif) conduisent à un
même modèle, la probabilité de celui-ci d'être le meilleur
modèle est accrue. L'exécution successive des deux procédés
30 constructif et destructif nécessite l'évaluation de P^2
modèles, c'est-à-dire un nombre très inférieur à 2^P modèles
nécessaires pour l'évaluation de la totalité des modèles.

Les procédés de sélection des modèles nécessitent donc
de nombreuses estimations de paramètres et l'emploi de tests
35 d'hypothèses statistiques ou de critères d'information qui
ne sont pas toujours faciles à comprendre par les utili-
sateurs non initiés.

L'invention met en oeuvre un nouveau procédé pour la construction du réseau de neurones dans lequel est utilisé un nouveau procédé d'évaluation des modèles. Plus précisément, selon l'invention, les descripteurs sont ordonnés par ordre de signification décroissante. Au départ, l'ensemble des P descripteurs est suffisamment important pour décrire les données. Parmi ces P descripteurs, on détermine celui qui décrit le mieux la sortie voulue, puis le second et ainsi de suite. On obtient ainsi un classement des descripteurs. On considère alors les sous-modèles constitués par un seul descripteur, deux descripteurs, trois descripteurs, etc., en commençant à chaque fois par le descripteur le plus significatif. Il est donc possible de considérer un nombre très réduit de modèles. En outre, selon l'invention, on utilise au moins une variable supplémentaire qui a un descripteur supplémentaire qui est aléatoire, c'est-à-dire que les valeurs de la variable supplémentaire sont purement aléatoires. Lorsque les descripteurs sont ordonnés, on considère que tous ceux qui se trouvent après le descripteur aléatoire ont une signification qui n'est pas supérieure à celle du descripteur aléatoire et peuvent donc être éliminés.

Plus précisément, dans un premier aspect, l'invention concerne un procédé de construction d'un réseau de neurones destiné à la modélisation d'un phénomène, le réseau comprenant des entrées destinées à recevoir des signaux représentatifs de valeurs de variables, des neurones destinés à appliquer une fonction d'activation aux signaux qu'ils reçoivent, au moins une sortie destinée à transmettre des données de résultat du modèle du phénomène, et des liaisons formées entre les entrées et les neurones et entre les neurones et la sortie, du type qui comprend, dans une première étape, la détermination des variables qui doivent être utilisées dans des modèles du phénomène par détermination de descripteurs représentatifs chacun des valeurs d'une variable, dans une seconde étape, la sélection des variables à incorporer à au moins un modèle optimal du

phénomène par évaluation des résultats de plusieurs modèles, et dans une troisième étape, la construction d'un réseau de neurones par détermination des liaisons des neurones en fonction d'un modèle optimal obtenu ; selon l'invention, le
5 procédé comprend aussi,

pendant ou avant la première étape de détermination des descripteurs, l'introduction d'au moins une variable supplémentaire qui possède des valeurs aléatoires, et la détermination d'un descripteur représentatif des valeurs de cette
10 variable supplémentaire,

le classement des descripteurs, y compris celui de la variable supplémentaire, par application d'un critère de comparaison des résultats donnés par les modèles aux données représentatives du résultat du phénomène, avec détermination
15 d'un ordre de signification décroissante des descripteurs, puis

l'élimination d'au moins un descripteur qui, dans l'ordre de signification décroissante des descripteurs, est classé après le descripteur représentatif des valeurs de la
20 variable supplémentaire.

Dans un mode de réalisation avantageux, le procédé comporte en outre la représentation des descripteurs et du résultat du phénomène par des vecteurs d'un espace à N dimensions, N étant le nombre d'exemples d'un ensemble
25 d'exemples d'apprentissage du phénomène, chaque exemple comprenant au moins une valeur de chacune des variables et au moins une donnée représentative du résultat du phénomène pour les valeurs correspondantes des variables. Dans ce mode de réalisation, le critère de comparaison utilisé pour le
30 classement des descripteurs est avantageusement une comparaison, dans l'espace à N dimensions, des angles formés par un vecteur représentatif d'un descripteur avec le vecteur représentatif du résultat du phénomène.

Dans ce mode de réalisation, l'étape de classement
35 comprend de préférence la détermination du premier descripteur dans l'ordre de signification décroissante des descripteurs, et la projection des vecteurs descripteurs

restants et du vecteur résultat sur l'espace à une dimension de moins qui est orthogonal à ce premier descripteur ; ensuite, cette étape comprend le classement des descripteurs dans cet espace à une dimension de moins pour la détermination du premier, dans l'ordre de signification décroissante, des descripteurs restants, et la projection des vecteurs descripteurs restants et du vecteur résultat sur un espace à une dimension de moins qui est orthogonal au premier descripteur dans l'ordre de signification décroissante des descripteurs restants, puis la répétition de ces étapes jusqu'au classement de tous les descripteurs ou jusqu'au classement du descripteur représentatif des valeurs de la variable supplémentaire.

De préférence, la construction d'au moins un modèle optimal du phénomène par évaluation des résultats de plusieurs modèles comprend la construction de plusieurs sous-modèles successifs du phénomène, chaque sous-modèle contenant une variable de plus que le sous-modèle précédent, la variable ajoutée étant choisie dans l'ordre de signification décroissante des descripteurs, la variable du premier sous-modèle étant soit une constante, soit la variable la plus significative, et la sélection d'un sous-modèle comme modèle optimal par utilisation d'un critère de sélection.

Dans cet exemple de réalisation, le critère de sélection d'un sous-modèle comprend de préférence la sélection du sous-modèle ayant le plus grand nombre de descripteurs donnant un niveau de risque de sélection de la variable supplémentaire qui est inférieur à un niveau choisi de seuil.

Dans un second aspect, l'invention concerne un procédé de construction d'un réseau de neurones destiné à la modélisation d'un phénomène, le réseau comprenant des entrées destinées à recevoir des signaux représentatifs de valeurs de variables qui sont représentées par des descripteurs, des neurones destinés à appliquer une fonction d'activation aux signaux qu'ils reçoivent, au moins une sortie destinée à transmettre des données de résultat du modèle du phénomène,

et des liaisons formées entre les entrées et les neurones et entre les neurones et la sortie, par détermination des liaisons des neurones en fonction du modèle ; le procédé comprend :

5 - la construction d'un réseau de neurones à une seule couche dont le nombre de neurones est certainement trop élevé, les entrées des neurones correspondant aux descripteurs du modèle, le réseau de neurones contenant en outre, dans sa couche unique, au moins un neurone supplémentaire
10 ayant une fonction d'activation dont les paramètres ont des valeurs aléatoires, et

 - l'exécution d'un processus comprenant, avec le nombre trop élevé de neurones, un apprentissage des neurones par utilisation des descripteurs, et l'élimination au moins du
15 neurone ayant la contribution la moins significative au résultat, afin que le réseau ait un nombre plus petit de neurones, puis

 - la répétition de ce processus avec le nombre plus petit de neurones, au moins jusqu'à ce que le neurone à
20 éliminer soit un neurone supplémentaire.

Dans ce mode de réalisation, l'apprentissage des neurones par utilisation des descripteurs est effectué de préférence avec une partie seulement des exemples. Il est avantageux que l'exécution d'un processus comprenne, avant
25 l'élimination d'un neurone, au moins une répétition d'un apprentissage pour la confirmation du neurone ayant la contribution la moins significative.

Il est avantageux que le modèle du phénomène utilisé soit un modèle optimal obtenu par mise en oeuvre du procédé
30 selon le premier aspect de l'invention.

D'autres caractéristiques et avantages de l'invention ressortiront mieux de la description qui suit d'un exemple de réalisation, faite en référence au dessin annexé sur lequel :

35 - la figure 1 est un diagramme vectoriel représentant géométriquement un algorithme de comparaison des descripteurs ;

- la figure 2 est un graphique indiquant les résultats obtenus d'une part avec une procédure dite "de Gram-Schmidt" et d'autre part avec évaluation des performances de tous les sous-modèles d'un ensemble complet ;

5 - la figure 3 est un graphique indiquant le résultat donné par l'algorithme d'évaluation de Gram-Schmidt et la répartition du classement de la variable aléatoire ; et

10 - la figure 4 est un graphique illustrant un processus modélisé dans un exemple de mise en oeuvre du procédé de l'invention.

On considère un exemple plus détaillé de mise en oeuvre de l'invention en référence à un exemple dans lequel on cherche à modéliser un processus.

15 On dispose de P descripteurs (c'est-à-dire qu'on suppose initialement que P variables peuvent participer au résultat). On construit donc P descripteurs sous forme de vecteurs dans un espace à N dimensions, N étant le nombre d'exemples. Chaque exemple comprend une valeur de chacune des P variables, et au moins la valeur d'un résultat.

20 Pour le classement des descripteurs, on utilise avantageusement l'algorithme d'orthogonalisation de Gram-Schmidt modifié qu'on décrit maintenant rapidement. On peut cependant se reporter avantageusement, pour plus de détails, à l'article de S.Chen, S.A.Billings et W.Luo, "Orthogonal
25 least squares methods and their application to non-linear system identification", International Journal of Control, Vol. 50, n° 5, p. 1873 à 1896, 1989.

30 L'algorithme d'orthogonalisation de Gram-Schmidt considère les descripteurs et la sortie voulue comme des vecteurs. Les notations sont les suivantes :

$$X = \begin{bmatrix} x_1^1 & \dots & x_p^1 \\ \vdots & & \vdots \\ x_1^N & \dots & x_p^N \end{bmatrix} = [X_1 \dots X_p]$$

$$\text{avec } X_p = \begin{bmatrix} x_p^1 \\ \vdots \\ x_p^N \end{bmatrix} \text{ à l'entrée } P \text{ et } Y = \begin{bmatrix} y^1 \\ \vdots \\ y^N \end{bmatrix}$$

5

La matrice X est la matrice des entrées (P colonnes correspondent aux P descripteurs du modèle et N lignes représentent les N exemples de l'ensemble d'apprentissage). On considère que la matrice X est composée de P vecteurs
 10 représentant chacun une entrée. Le vecteur Y est le vecteur de sortie (N lignes correspondent aux sorties observées des N exemples).

A la première itération, on détermine le vecteur d'entrée qui "explique" le mieux la sortie. Pour cela, on
 15 détermine l'angle du vecteur de sortie avec chaque vecteur d'entrée. On évalue à cet effet le carré des cosinus des angles. Le vecteur sélectionné est celui pour lequel le carré des cosinus est maximal.

Une fois déterminé ce vecteur le plus significatif, on
 20 élimine sa contribution en projetant le vecteur de sortie et tous les vecteurs d'entrée restants sur un sous-espace ou espace à N-1 dimensions qui est orthogonal au vecteur sélectionné.

L'algorithme se poursuit jusqu'à ce que tous les
 25 vecteurs d'entrée aient été ordonnés. Selon l'invention, on peut interrompre l'évaluation lorsqu'on doit sélectionner le vecteur aléatoire.

A chaque itération, on calcule la solution ordinaire des moindres carrés et la valeur de l'écart quadratique
 30 moyen correspondant. L'estimation des paramètres de la régression des moindres carrés est obtenue par résolution d'une équation linéaire ayant une matrice triangulaire supérieure et la norme du vecteur de sortie projeté détermine la valeur de l'écart quadratique moyen.

35 La figure 1 indique l'interprétation géométrique de l'algorithme qu'on vient de décrire. Sur cette figure, on a représenté un espace à deux dimensions. Le vecteur de sortie

Y est mieux "expliqué" par le vecteur X_2 que par le vecteur X_1 (l'angle θ_2 est plus petit que l'angle θ_1). On sélectionne donc X_2 comme premier descripteur. Pour éliminer la partie expliquée par ce descripteur, on projette les vecteurs Y et X_1 (et de façon générale tous les vecteurs restants) sur le sous-espace orthogonal au vecteur X_2 . Les projections sont utilisées pour la sélection du descripteur suivant mais, dans le cas de deux dimensions, il n'en existe plus puisqu'il ne reste plus qu'un seul vecteur d'entrée X_1 .

10 L'algorithme de Gram-Schmidt qu'on vient de décrire ne donne pas toujours le résultat optimal. La figure 2 indique les résultats obtenus d'une part avec l'algorithme de Gram-Schmidt et d'autre part avec évaluation des performances des 1 024 sous-modèles d'un ensemble complet comprenant quinze points d'apprentissage de dix descripteurs dont cinq seulement sont pertinents. Les croix représentent les résultats des 1 024 sous-modèles possibles et la courbe le sous-modèle sélectionné par l'algorithme de Gram-Schmidt. On note que, à l'exception du sous-modèle à trois descripteurs, les sous-modèles obtenus sont toujours les meilleurs.

20 On a porté sur la figure 3 en superposition, en fonction du nombre de descripteurs, le résultat donné par l'algorithme d'évaluation de Gram-Schmidt et la répartition du classement de la variable aléatoire, avec indication, sur l'échelle des ordonnées de droite, de la probabilité en pourcentage. On note ainsi que la probabilité pour que la variable aléatoire soit comprise dans les cinq premiers descripteurs est inférieure à 10 %. On peut ainsi déterminer que, si l'on sélectionne un sous-modèle à cinq descripteurs, la probabilité pour qu'une variable aléatoire explique mieux le problème posé qu'un des cinq descripteurs sélectionnés est inférieure à 10 %. Le niveau de risque détermine le nombre de descripteurs retenu. Ce niveau de risque ne doit pas être trop élevé, car des variables non significatives peuvent alors d'être incorporées. Il ne doit pas être trop faible car des valeurs significatives peuvent de ne pas être incorporées. Dans le cas représenté, les seules possibilités

de sélection sont cinq ou six descripteurs, c'est-à-dire le nombre réel de descripteurs significatifs ou ce nombre augmenté d'un descripteur non significatif.

5 Cette répartition du classement de la variable aléatoire peut aussi être réalisée uniquement par le calcul, mais on ne décrit pas celui-ci.

10 Le traitement qu'on vient de décrire permet ainsi de déterminer les descripteurs qui doivent être conservés et le modèle optimal. On peut alors construire un réseau de neurones.

15 On a déjà démontré que n'importe quel réseau de neurones à plusieurs couches de type non bouclé pouvait être représenté par un réseau de neurones à une seule couche cachée. On utilise donc initialement un réseau de neurones à une couche cachée dont le nombre de descripteurs (couche d'entrées) a été déterminé, et ayant un nombre de neurones trop élevé, puis on élimine les neurones qui n'ont pas une contribution significative. On poursuit l'apprentissage avec les neurones restants, et on élimine à nouveau les neurones
20 inutiles. On arrête la procédure lorsqu'on n'élimine plus aucun neurone.

Dans un mode de réalisation particulièrement avantageux de l'invention, on utilise un processus analogue à celui de la sélection des descripteurs pour la sélection des neu-
25 rones. Plus précisément, on introduit un neurone supplémentaire ayant une fonction d'activation qui n'est pas linéaire et dont les paramètres sont aléatoires. Dans cette réalisation, on exécute la procédure jusqu'à ce que ce neurone supplémentaire se classe après les autres neurones.
30 (De manière connue, les fonctions d'activation sont continues, dérivables et bornées, et des exemples sont les fonctions trigonométriques hyperboliques, telles que la tangente hyperbolique, et les fonctions gaussiennes).

35 Si l'on dispose d'un très grand nombre d'exemples pour l'apprentissage, il est possible que le neurone supplémentaire soit tout de suite classé le dernier. Dans ce cas, l'utilisation d'un tel neurone supplémentaire ne présente

pas d'intérêt. Il est alors préférable d'utiliser un sous-ensemble réduit pour l'apprentissage, afin que le neurone supplémentaire ne soit pas immédiatement le dernier. L'apprentissage s'effectue sur les exemples de ce sous-ensemble, on conserve la valeur moyenne de l'écart quadratique sur le reste de l'ensemble, et on applique la procédure de sélection aux exemples du sous-ensemble ; les coefficients du réseau de neurones correspondent à la valeur moyenne minimale de l'écart quadratique ainsi calculé.

De cette manière, on supprime les neurones classés après le neurone supplémentaire.

Le procédé selon l'invention présente, grâce au classement des descripteurs, l'intérêt d'indiquer quelles sont les variables les plus significatives. Il permet une réduction considérable du temps de calcul nécessaire pour l'évaluation des descripteurs significatifs, puis pour la construction du réseau de neurones.

L'invention concerne aussi des réseaux de neurones réalisés par mise en oeuvre du procédé précité. Ces réseaux de neurones, lorsque leur structure optimale a été ainsi évaluée par mise en oeuvre du procédé de l'invention, peuvent être réalisés par exemple sous forme de circuits intégrés, avec détermination des liaisons entre les entrées, les neurones et la sortie ou les sorties, et avec détermination des fonctions d'activation des neurones.

Exemple

On considère maintenant, à titre d'illustration, un exemple d'application de l'invention à la solution d'un problème de modélisation destiné à la simulation d'un processus.

La figure 4 est un graphique représentant, en ordonnées, la valeur donnée par un processus (sur une échelle allant de -15 à +15) en fonction du temps t , porté en abscisses. La courbe en trait gras représente la valeur donnée $y_p(t)$ par le processus en fonction d'une commande $u(t)$ représentée par la courbe en trait fin.

Dans la première phase du procédé, on choisit 20 variables possibles $y_p(t-1)$ à $y_p(t-10)$ et $u(t-1)$ à $u(t-10)$. Le graphique de la figure 4 permet d'établir 20 descripteurs correspondant aux 20 variables pour 1 000 exemples. On
5 ajoute un descripteur aléatoire, on exécute la première phase du procédé, et on obtient finalement les 3 variables $y_p(t-1)$, $y_p(t-2)$ et $u(t-1)$.

On construit alors un réseau de neurones représentatif de ce processus. On utilise initialement un réseau à 20
10 neurones, plus un neurone aléatoire, chaque neurone ayant une fonction d'activation en forme de sigmoïde. Après un premier passage, il reste 17 neurones. Après un second passage, il reste 14 neurones. Le traitement s'arrête à 11 ou 12 neurones.

15 Pour évaluer l'intérêt du procédé de l'invention, on construit 21 réseaux de neurones différents (de 0 à 20 neurones), et on les compare pour déterminer le meilleur, par détermination de l'écart quadratique moyen en fonction du nombre de neurones. Cette évaluation est très longue et
20 nécessite d'importants moyens de calcul. Le résultat indique que le meilleur réseau comporte 11 neurones. Ce résultat confirme l'exactitude du résultat obtenu beaucoup plus rapidement par le procédé de l'invention.

Il est bien entendu que l'invention n'a été décrite et
25 représentée qu'à titre d'exemple préférentiel et qu'on pourra apporter toute équivalence technique dans ses éléments constitutifs sans pour autant sortir de son cadre.

REVENDICATIONS

1. Procédé de construction d'un réseau de neurones destiné à la modélisation d'un phénomène, le réseau comprenant des entrées destinées à recevoir des signaux représentatifs de valeurs de variables, des neurones destinés à
5 appliquer une fonction d'activation aux signaux qu'ils reçoivent, au moins une sortie destinée à transmettre des données de résultat du modèle du phénomène, et des liaisons formées entre les entrées et les neurones et entre les
10 neurones et la sortie,

du type qui comprend :

dans une première étape, la détermination des variables qui doivent être utilisées dans des modèles du phénomène par détermination de descripteurs représentatifs chacun des
15 valeurs d'une variable,

dans une seconde étape, la sélection des variables à incorporer à au moins un modèle optimal du phénomène par évaluation des résultats de plusieurs modèles, et

dans une troisième étape, la construction d'un réseau
20 de neurones par détermination des liaisons des neurones en fonction d'un modèle optimal obtenu,

caractérisé en ce que le procédé comprend

- pendant ou avant la première étape de détermination des descripteurs, l'introduction d'au moins une variable
25 supplémentaire qui possède des valeurs aléatoires, et la détermination d'un descripteur représentatif des valeurs de cette variable supplémentaire,

- le classement des descripteurs, y compris celui de la variable supplémentaire, par application d'un critère de
30 comparaison des résultats donnés par les modèles aux données représentatives du résultat du phénomène, avec détermination d'un ordre de signification décroissante des descripteurs, puis

- l'élimination d'au moins un descripteur qui, dans
35 l'ordre de signification décroissante des descripteurs, est classé après le descripteur représentatif des valeurs de la variable supplémentaire.

2. Procédé selon la revendication 1, caractérisé en ce qu'il comporte en outre la représentation des descripteurs et du résultat du phénomène par des vecteurs d'un espace à N dimensions, N étant le nombre d'exemples d'un ensemble d'exemples d'apprentissage du phénomène, chaque exemple comprenant au moins une valeur de chacune des variables et au moins une donnée représentative du résultat du phénomène pour les valeurs correspondantes des variables.

3. Procédé selon la revendication 2, caractérisé en ce que le critère de comparaison utilisé pour le classement des descripteurs est une comparaison, dans l'espace à N dimensions, des angles formés par un vecteur représentatif d'un descripteur avec le vecteur représentatif du résultat du phénomène.

4. Procédé selon la revendication 3, caractérisé en ce que l'étape de classement comprend la détermination du premier descripteur dans l'ordre de signification décroissante des descripteurs, et la projection des vecteurs descripteurs restants et du vecteur résultat sur l'espace à une dimension de moins qui est orthogonal à ce premier descripteur, puis le classement des descripteurs dans cet espace à une dimension de moins pour la détermination du premier, dans l'ordre de signification décroissante, des descripteurs restants, et la projection des vecteurs descripteurs restants et du vecteur résultat sur un espace à une dimension de moins qui est orthogonal au premier descripteur dans l'ordre de signification décroissante des descripteurs restants, et la répétition de ces étapes jusqu'au classement de tous les descripteurs ou jusqu'au classement du descripteur représentatif des valeurs de la variable supplémentaire.

5. Procédé selon l'une quelconque des revendications précédentes, caractérisé en ce que la construction d'au moins un modèle optimal du phénomène par évaluation des résultats de plusieurs modèles, comprend

- la construction de plusieurs sous-modèles successifs du phénomène, chaque sous-modèle contenant une variable de

plus que le sous-modèle précédent, la variable ajoutée étant choisie dans l'ordre de signification décroissante des descripteurs, la variable du premier sous-modèle étant soit une constante, soit la variable la plus significative, et

- 5 - la sélection d'un sous-modèle comme modèle optimal par utilisation d'un critère de sélection.

6. Procédé selon la revendication 5, caractérisé en ce que le critère de sélection d'un sous-modèle comprend la sélection du sous-modèle ayant le plus grand nombre de descripteurs donnant un niveau de risque de sélection de la variable supplémentaire qui est inférieur à un niveau choisi de seuil.

7. Procédé de construction d'un réseau de neurones destiné à la modélisation d'un phénomène, le réseau comprenant des entrées destinées à recevoir des signaux représentatifs de valeurs de variables qui sont représentées par des descripteurs, des neurones destinés à appliquer une fonction d'activation aux signaux qu'ils reçoivent, au moins une sortie destinée à transmettre des données de résultat du modèle du phénomène, et des liaisons formées entre les entrées et les neurones et entre les neurones et la sortie, par détermination des liaisons des neurones en fonction du modèle, caractérisé en ce qu'il comprend :

- 25 - la construction d'un réseau de neurones à une seule couche dont le nombre de neurones est certainement trop élevé, les entrées des neurones correspondant aux descripteurs du modèle, le réseau de neurones contenant en outre, dans sa couche unique, au moins un neurone supplémentaire ayant une fonction d'activation dont les paramètres ont des valeurs aléatoires, et

- 30 - l'exécution d'un processus comprenant, avec le nombre trop élevé de neurones, un apprentissage des neurones par utilisation des descripteurs, et l'élimination au moins du neurone ayant la contribution la moins significative au résultat, afin que le réseau ait un nombre plus petit de neurones, puis

- la répétition de ce processus avec le nombre plus petit de neurones, au moins jusqu'à ce que le neurone à éliminer soit un neurone supplémentaire.

5 8. Procédé selon la revendication 7, caractérisé en ce que l'apprentissage des neurones par utilisation des descripteurs est effectué avec une partie seulement des exemples.

10 9. Procédé selon l'une des revendications 7 et 8, caractérisé en ce que l'exécution d'un processus comprend, avant l'élimination d'un neurone, au moins une répétition d'un apprentissage pour la confirmation du neurone ayant la contribution la moins significative.

15 10. Procédé selon l'une quelconque des revendications 7 à 9, caractérisé en ce que le modèle du phénomène utilisé est un modèle optimal obtenu par mise en oeuvre d'un procédé selon l'une quelconque des revendications 1 à 6.

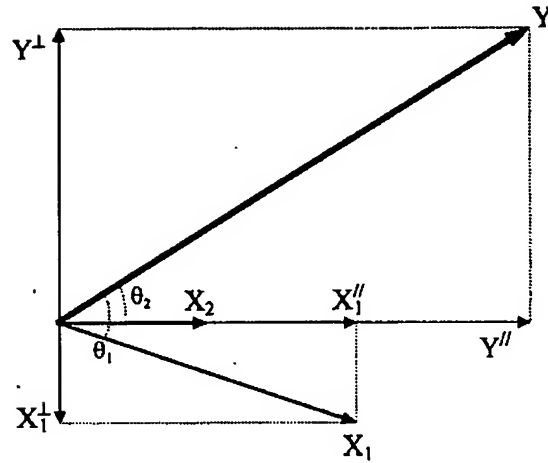


Fig. 1

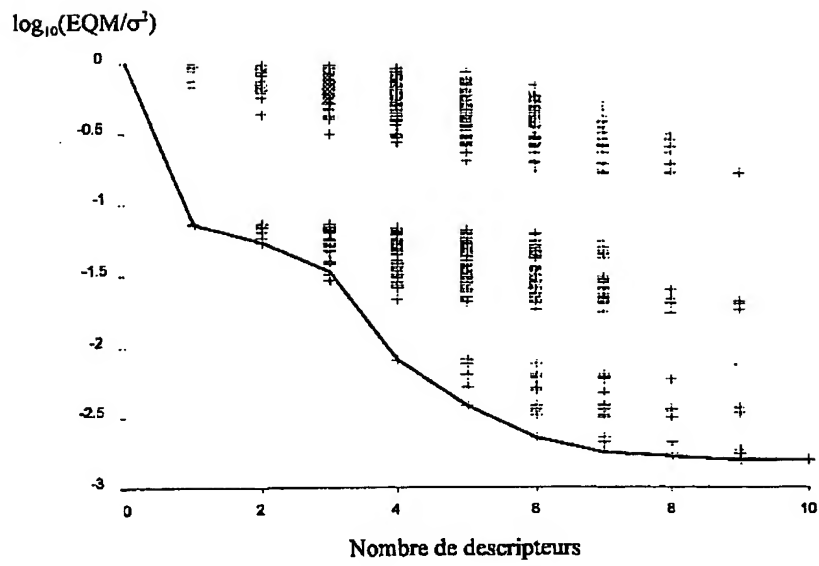


Fig. 2

2 / 2

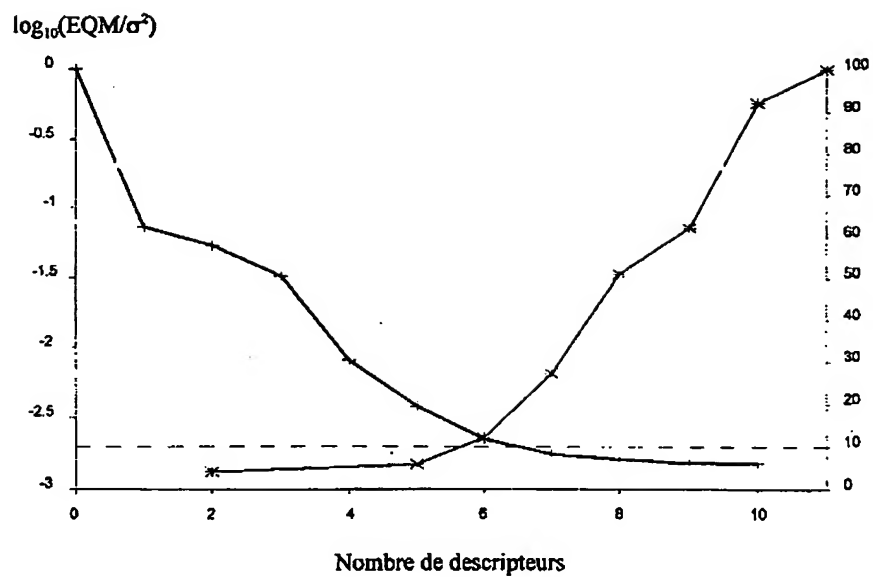


Fig. 3

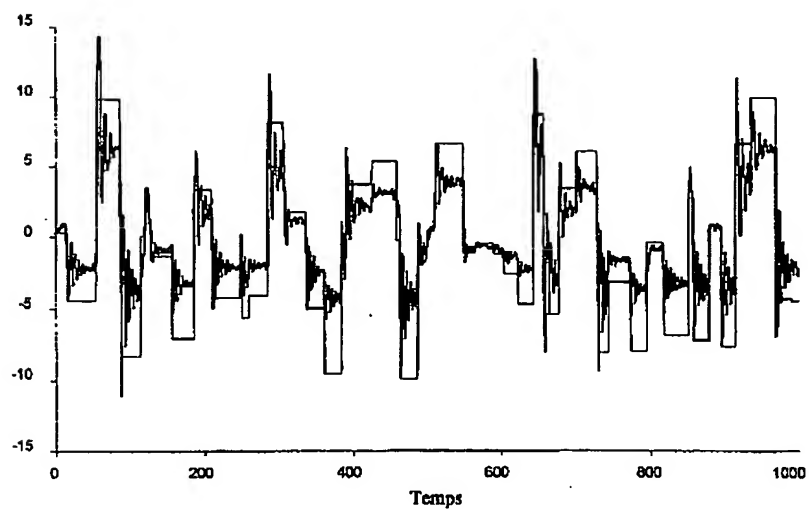


Fig. 4

INTERNATIONAL SEARCH REPORT

Intern. Application No.
PCT/FR 98/01381

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06F15/80

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 94 25933 A (PAVILION TECH INC) 10 November 1994 see page 4, line 2 - line 27	1,7
A	PARBERRY I ET AL: "RELATING BOLTZMANN MACHINES TO CONVENTIONAL MODELS OF COMPUTATION" NEURAL NETWORKS, vol. 2, no. 1, 1 January 1989, pages 59-67, XP000084477 see page 63, left-hand column, line 3 - page 65, left-hand column, line 19; figures 3-7	1,7
A	EP 0 528 399 A (TOYODA MACHINE WORKS LTD) 24 February 1993 see page 3, line 9 - page 4, line 55 -/--	1,7

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- "&" document member of the same patent family

Date of the actual completion of the international search

8 September 1998

Date of mailing of the international search report

16/09/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel (+31-70) 340-2040, Tx. 31 651 epo rd,
Fax (+31-70) 340-3016

Authorized officer

Schenkels, P

INTERNATIONAL SEARCH REPORT

Intern of Application No
PCT/FR 98/01381

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>WALEED FAKHR ET AL: "MINIMUM DESCRIPTION LENGTH PRUNING AND MAXIMUM MUTUAL INFORMATION TRAINING OF ADAPTIVE PROBABILISTIC NEURAL NETWORKS" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS (ICN, SAN FRANCISCO, MAR. 28 - APR. 1, 1993, vol. VOL. 3, no. -, 28 March 1993, pages 1338-1342, XP000379471 INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS see page 1338, left-hand column, line 1 - page 1340, left-hand column, line 2</p>	1,7
A	<p>GALLANT S I ET AL: "Random cells: an idea whose time has come and gone... and come again?" IEEE FIRST INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, SAN DIEGO, CA, USA, 21-24 JUNE 1987, pages 671-678 vol.2, XP000091982 1987, SAN DIEGO, CA, USA, SOS PRINTING, USA see page 671, line 1 - page 675, line 16; figures 1-4</p>	1,7

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/FR 98/01381

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9425933 A	10-11-1994	AU 6669594 A	21-11-1994
		CA 2161655 A	10-11-1994
		EP 0696372 A	14-02-1996
EP 0528399 A	24-02-1993	JP 5046585 A	26-02-1993
		JP 5054014 A	05-03-1993
		US 5355434 A	11-10-1994

RAPPORT DE RECHERCHE INTERNATIONALE

Dema internationale No
PCT/FR 98/01381

A. CLASSEMENT DE L'OBJET DE LA DEMANDE
CIB 6 G06F15/80

Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB

B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE

Documentation minimale consultée (système de classification suivi des symboles de classement)
CIB 6 G06F

Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche

Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si c'est réalisable, termes de recherche utilisés)

C. DOCUMENTS CONSIDERES COMME PERTINENTS

Catégorie *	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	WO 94 25933 A (PAVILION TECH INC) 10 novembre 1994 voir page 4, ligne 2 - ligne 27 ---	1,7
A	PARBERRY I ET AL: "RELATING BOLTZMANN MACHINES TO CONVENTIONAL MODELS OF COMPUTATION" NEURAL NETWORKS, vol. 2, no. 1, 1 janvier 1989, pages 59-67, XP000084477 voir page 63, colonne de gauche, ligne 3 - page 65, colonne de gauche, ligne 19; figures 3-7 ---	1,7
A	EP 0 528 399 A (TOYODA MACHINE WORKS LTD) 24 février 1993 voir page 3, ligne 9 - page 4, ligne 55 --- -/-	1,7

☒ Voir la suite du cadre C pour la fin de la liste des documents

☒ Les documents de familles de brevets sont indiqués en annexe

* Catégories spéciales de documents cités:

- "A" document définissant l'état général de la technique, non considéré comme particulièrement pertinent
- "E" document antérieur, mais publié à la date de dépôt international ou après cette date
- "L" document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (celle qu'indiquée)
- "O" document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens
- "P" document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée

"T" document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention

"X" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément

"Y" document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier

"Z" document qui fait partie de la même famille de brevets

Date à laquelle la recherche internationale a été effectivement achevée

8 septembre 1998

Date d'expédition du présent rapport de recherche internationale

16/09/1998

Nom et adresse postale de l'administration chargée de la recherche internationale
Office Européen des Brevets, P.B. 5818 Patentaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3018

Fonctionnaire autorisé

Schenkels, P

RAPPORT DE RECHERCHE INTERNATIONALE

Demar internationale No
PCT/FR 98/01381

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	<p>WALEED FAKHR ET AL: "MINIMUM DESCRIPTION LENGTH PRUNING AND MAXIMUM MUTUAL INFORMATION TRAINING OF ADAPTIVE PROBABILISTIC NEURAL NETWORKS" PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS (ICN, SAN FRANCISCO, MAR. 28 - APR. 1, 1993, vol. VOL. 3, no. -, 28 mars 1993, pages 1338-1342, XP000379471 INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS voir page 1338, colonne de gauche, ligne 1 - page 1340, colonne de gauche, ligne 2 -----</p>	1,7
A	<p>GALLANT S I ET AL: "Random cells: an idea whose time has come and gone... and come again?" IEEE FIRST INTERNATIONAL CONFERENCE ON NEURAL NETWORKS, SAN DIEGO, CA, USA, 21-24 JUNE 1987, pages 671-678 vol.2, XP000091982 1987, SAN DIEGO, CA, USA, SOS PRINTING, USA voir page 671, ligne 1 - page 675, ligne 16; figures 1-4 -----</p>	1,7

RAPPORT DE RECHERCHE INTERNATIONALE

Renseignements relatifs au ...nombres de familles de brevets

Demande internationale No

PCT/FR 98/01381

Document brevet cité au rapport de recherche	Date de publication	Membre(s) de la famille de brevet(s)	Date de publication
WO 9425933 A	10-11-1994	AU 6669594 A	21-11-1994
		CA 2161655 A	10-11-1994
		EP 0696372 A	14-02-1996
EP 0528399 A	24-02-1993	JP 5046585 A	26-02-1993
		JP 5054014 A	05-03-1993
		US 5355434 A	11-10-1994